

1968

# The Temporal Reliability of the Length-Complexity Index

Margaret C. Barlow  
*Eastern Illinois University*

---

## Recommended Citation

Barlow, Margaret C., "The Temporal Reliability of the Length-Complexity Index" (1968). *Masters Theses*. 4158.  
<https://thekeep.eiu.edu/theses/4158>

This is brought to you for free and open access by the Student Theses & Publications at The Keep. It has been accepted for inclusion in Masters Theses by an authorized administrator of The Keep. For more information, please contact [tabruns@eiu.edu](mailto:tabruns@eiu.edu).

PAPER CERTIFICATE #3

To: Graduate Degree Candidates who have written formal theses.

Subject: Permission to reproduce theses.

The University Library is receiving a number of requests from other institutions asking permission to reproduce dissertations for inclusion in their library holdings. Although no copyright laws are involved, we feel that professional courtesy demands that permission be obtained from the author before we allow theses to be copied.

Please sign one of the following statements.

Booth Library of Eastern Illinois University has my permission to lend my thesis to a reputable college or university for the purpose of copying it for inclusion in that institution's library or research holdings.

5/20/68

Date

\_\_\_\_\_  
Author

I respectfully request Booth Library of Eastern Illinois University not allow my thesis be reproduced because \_\_\_\_\_

\_\_\_\_\_  
Date

\_\_\_\_\_  
Author

/LB1861.C57X8258>C2/

THE TEMPORAL RELIABILITY OF THE

LENGTH-COMPLEXITY INDEX

(TITLE)

BY

Margaret C. Barlow  
      

THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS  
FOR THE DEGREE OF

Master of Science

IN THE GRADUATE SCHOOL, EASTERN ILLINOIS UNIVERSITY  
CHARLESTON, ILLINOIS

1968

YEAR

I HEREBY RECOMMEND THIS THESIS BE ACCEPTED AS FULFILLING  
THIS PART OF THE GRADUATE DEGREE CITED ABOVE

May 20, 1968  
DATE

May 20, 1968  
DATE

#### ACKNOWLEDGEMENTS

The writer wishes to express her profound appreciation to those who have given advice and guidance in the preparation of this thesis.

Gratitude and recognition are extended to Mr. Lynn E. Miner for his diligent direction of this thesis. Acknowledgement is extended to Dr. Wayne L. Thurman and Dr. Jerry Griffith who constantly motivated and inspired the author to broaden her understanding and knowledge of her professional field. The author is conscious of her debt to the faculty and graduate students at Eastern Illinois University for their help and encouragement. The author wishes to thank the administration of the Sullivan Public Schools for their cooperation in this study. Finally, the author is grateful to her husband and family for their support and cooperation in this project.

## TABLE OF CONTENTS

<b>Chapter</b>	
I. STATEMENT OF THE PROBLEM. . . . .	1
Introduction . . . . .	1
Definitions . . . . .	5
II. REVIEW OF THE LITERATURE . . . . .	7
III. SUBJECTS, PROCEDURES, EQUIPMENT . . . . .	13
Selection of Subjects . . . . .	13
Procedures . . . . .	16
IV. RESULTS AND DISCUSSION . . . . .	20
Results . . . . .	20
Discussion . . . . .	23
V. SUMMARY AND CONCLUSIONS . . . . .	27
Summary . . . . .	27
Conclusions . . . . .	31
APPENDIX . . . . .	33
BIBLIOGRAPHY . . . . .	40

# LIST OF TABLES

Table	Page
1. Intraclass Correlation Coefficients for MLR and LCI . . . . .	20
2. MLR and LCI Scores for Each Sample Time . . . . .	21
3. Within and Between Correlation Coefficients for MLR and LCI . . . . .	22

## CHAPTER I

### STATEMENT OF THE PROBLEM

#### Introduction

The origin of the language known today as English finds its roots thousands of years ago in the family of Indo-European languages. Laird, in The Miracle of Language, says,

We know only that there must have been a time when there was no language, and then there was a time when there was a language, but we do not know how, when, where, or by whom language came into being. (p. 23)

Some general ideas about language can be briefly summarized. Language represents the perpetuation of mankind; it is man's way of achieving immortality. Language is a system enabling a man to define and organize his environment. Language is the only means of preserving the past--and the future is also dependent upon it. However, language is different as viewed by different disciplines.

The discipline of lexicography does not know exactly how large the English vocabulary is, with its synonyms and technical terms for each field. It may be the largest, most precise, ever available to the speakers of a single language. The basic Indo-European vocabulary was small, but speakers of English have borrowed prodigiously from other languages as well.

The discipline of semantics views words as symbols or carriers of meaning. Therefore many words which are single in form, become multiple

in meaning. There may be nearly twice as many meanings in English as there are recognized words.

Generative grammarians are presently trying to discover the rules by which children formulate their utterances. They have rejected the rules of traditional English grammar because it was derived from classical written Latin, which was basically an inflected language. In Latin, change in the form of a word, or a change in pitch, or the addition of an ending determined its meaning in a sentence. In contrast to Latin, English is a distributive language. That is, the arrangement of the symbols, or syntax is one of the most important determinants of semantics. However, it is not the sole determinant.

Early descriptive efforts in structural linguistics were concentrated first on phonology, then later, on syntax. Psychologists usually studied language as it related to inner cognitive skills. Chomsky, by providing a new generative conception of grammar showed how syntax could provide a common ground for the collaboration of linguists and psychologists. (Osgood and Sebeok, 1965, p. 213)

There is a distinction between competence and performance in children's language which is a fundamental one for Chomsky (1965).

"Among linguists, the difference between competence and performance takes the form of distinguishing between what a person knows about a language and his expression of this knowledge in talking and listening." (McNeill, 1966) Grammar is developed to describe the child's linguistic competence. By generative, Chomsky (1957) implies a system of rules that in some explicit and well-defined way assigns structural descriptions to sentences.



Chomsky (1957) is concerned, not with a specific grammar, but with a general form that any language may assume. Perhaps the most difficult part to understand is that there are similarities which hold true for human languages everywhere. He calls these similarities "language universals." These universals have both linguistic and psychological dimensions. They need not be explained for any specific grammar for they are the same for all grammars. It is for this reason that Chomsky assumes that these universals, whatever they prove to be, do not need to be learned by a child, but represent an innate capacity (in Saporta, 1961, p. 37).

David McNeill adopted the Chomsky approach, but he has been concerned with the formulation of a theory to aid in understanding the facts of language acquisition. For McNeill (1966), "The fundamental problem to which we address ourselves in language acquisition by normal children is the simple fact that the process occurs in a surprisingly short period of time. Grammatical speech does not begin before 1.5 years of age; yet, as far as we can tell, acquisition is virtually complete by 3.5 or 4 years." He questions how such intricate competence can emerge in the short span of twenty-four to thirty months.

McNeill (1966) believes it is possible to describe performance without explaining it. If we wish to explain performance, we must show how it derives from competence. Those in language pathology, continue to be faced with the problem of sufficiently testing the performance of a child's language and inferring something about the competence. This distinction becomes important in differential diagnosis, because the rules of generative grammar reside in competence, not performance (McNeill, 1966).

If we are to make important judgments about the verbal maturity of children, we need to have clear-cut tools in our inferential system to relate the rules for performance to the rules for competence. In other words, it is important clinically to have accurate, sensitive measures of children's linguistic performance. Historically, mean length of response (MLR) has been the measure used most frequently. McCarthy (1930) developed a set of rules for identifying a response, counting the words in each response, and classifying each response with regard to grammatical completeness and complexity. Most researchers have used this measure with minor modifications until very recently.

Another approach to the study of children's verbal maturity has discussed the emergence of two and three word sentences in terms of certain "open" and "pivot" class constructions (Brown and Bellugi, 1964; Brown and Fraser, 1963; Brown, Fraser and Bellugi, 1963; Ervin, 1964; Miller and Ervin, 1964; Braine, 1963; McNeill, 1966a, 1966b). It examines the way in which children initially combine words into sentences. Another method proposes a tentative hierarchy of developmental sentence types. It proceeds from various kinds of two word combinations through noun phrase constructions and kernel sentences up to emerging transformations (Lee, 1966).

The newest classification system attempts to measure both the length and complexity (LCI) of children's utterances (Shriner, 1967a). It consists essentially of a numeric weighing scale for assessing developmental changes in child language. Each of the last three techniques attempts to describe the grammatical rules that children

employ in generating sentences (Miner, forthcoming). By rigorously describing the language of children, we can hope to quantify the developmental changes that occur in that language.

Any measurement tool must meet the criteria of validity and reliability. It is extremely difficult to establish the validity of a measure, but reliability can be measured four different ways: internal consistency, adequacy of item sampling, examiner reliability and temporal reliability.

In order for the LCI to be considered a reliable measure, it is necessary that several language samples taken over a short period of time should yield similar results. This is known as test-retest or temporal reliability. Replicating the Minifie, Darley and Sherman (1963) procedure, the purpose of this investigation is to determine the day-to-day consistency of the verbal output of children as measured by the LCI.

#### Definitions

The following terms from this study are defined for further clarification:

Mean length of response.--is usually defined as the number of words per response averaged over a sample of 50 responses. Most investigators have recorded one sample consisting of 50 responses from each child. Customarily, the responses are tape recorded and usually are elicited by pictures and/or toys presented by the examiner. The individual responses of each child are then transcribed by tape replay which constitutes a sample of language for that particular child. An MLR score is derived from

the sample and used for individual comparisons or pooled for group analysis.

Length-Complexity Index.--is a measure designed to make a composite analysis of sentence length and sentence complexity. It was synthesized by Shriner (1967a), and based on the research of Menyuk (1964), Bellugi (1964), and Cazden (1965). Menyuk noted that sentence complexity relates to more than sentence length; it is also a function of the ability to apply increasingly differentiated rules for generating sentences. Briefly, in the noun phrase part of the index, the child is given points for the use of modifiers, articles, plurals and possessive inflections. The child's final score contains a total of his noun phrase points plus his verb phrase points plus additional points for each sentence divided by the number of sentences.

Temporal reliability.--Test-retest reliability. This means that several language samples taken in close temporal proximity should yield similar results.

## CHAPTER II

### REVIEW OF THE LITERATURE

Studies of the way in which children's language develops have usually included some description of connected speech samples in terms of amount of verbal output and grammatical complexity of sentences used. In one of the earliest, and one of the most important studies of children's language, McCarthy (1930) elicited fifty consecutive verbal responses from children and manually recorded them, using pictures and toys as stimulus material. McCarthy developed a set of rules for identifying a response, counting the words in each response, and classifying each response with regard to grammatical completeness and complexity. She used mean length of response as her main measure of children's linguistic achievement. McCarthy originally called MLR the "simplest and most objective measure of the degree to which children combine words at the various ages;" more recently she has stated that, "no measure seems to have superseded the mean length of sentence for a reliable, easily determined, objective, quantitative, and easily understood measure of linguistic maturity." (1954).

Day (1932) used McCarthy's procedures, definitions, classifications and methods of analysis in her study of 2-5 year old twins, recording manually 50 consecutive verbal responses. She found MLR varies if the child is a twin.



Davis (1937) used essentially the same procedure as McCarthy in her study of twins, singletons with siblings, and only children. Davis used 436 children at three age levels, 5.5, 6.5 and 9.5. She also collected samples of 50 responses, usually consecutive, and for her analysis modified and clarified McCarthy's rules for sentence classification.

The most comprehensive study of children's language to date is that of Templin (1957), who compiled normative data concerning the speech and language development of children between the ages of 3.0 and 8.0 years. Among the language variables she investigated are articulation, discrimination, vocabulary, and verbalization. Templin also devised a quantitative method of representing sentence completeness-complexity. She assigned weights to the categories of the McCarthy-Davis outline in order to obtain a structural complexity score (SCS).

In addition to the studies of children's language mentioned above, Shriner (1968, in press) has compiled an extensive review of Mean Length of Response as a measure of expressive language development in children. He reports that MLR has been shown to vary (1) with CA and IQ (Fisher, 1934); (2) in conversation with an adult (Smith, 1935); (3) in conversation with peers (Smith, 1935; Hahn, 1948); (4) in classroom situations (Hahn, 1948); (5) with socioeconomic status (Templin, 1957); and with (6) stimulus materials and experimenter (Cowan, et al., 1964).

In order for the MLR to be considered a sufficient index of verbal maturity in children, it must satisfy the criteria of validity and reliability. Shriner (1968) says,

Since the validity of any test is difficult or practically impossible to prove directly, earlier investigators have placed added importance on the reliability of MLR. (Reliability is a necessary component of validity, but high reliability does not necessarily permit the conclusion that a sampling procedure is valid). From 50 response samples correlation coefficients have been recorded for assessing reliability for various purposes; that is, to evaluate and to estimate the degree of agreement between examiners in preparing transcripts (Siegel, 1962); (b) an examiner's repeated measures obtained from the transcripts (McCarthy, 1930; Day, 1932; Williams, 1937; Davis, 1937; Spriestersbach, Darley, and Morris, 1958; Winitz, 1959; Minifie, Darley and Sherman, 1963), and (c) examiners in obtaining measures from the same transcripts (Day, 1932; Davis, 1937; Spriestersbach et al., 1958, Winitz, 1959; Miller, 1961; Minifie, et al., 1963; and others), and (d) the same measures obtained in the same manner over a period of time from the same children (Fisher, 1934; Minifie et al., 1963).

Intra- and inter-examiner reliability coefficients from the language transcripts are in general agreement and show relatively high correlations (.90 and above). This indicates that children's utterances can be recorded and analyzed reliably for a single sample of 50 responses. It does not, however, indicate to what extent the 50 response samples represent the children's language development in general; that is, whether or not 'the items constituting the test adequately represent the entire universe of items which the test undertakes to sample' (Ansstasi, 1954, p. 97).

Darley and Moll (1960) were concerned specifically with the size of language sample needed to obtain an adequate reliability coefficient for MLR. They summarized previous research methodology by noting that some investigators do not specify their choice of sample size, while others have used different numbers of sentences. Darley and Moll concluded from their reliability analysis that a 50-response sample would have an estimated reliability coefficient of .85, adequate for most purposes; however, this would depend on the precision needed by an examiner in a particular situation. Theoretically, the larger the sample and the greater the reduction in variability, the more closely the measure approaches the

child's "true" mean. The time required, however, to record, transcribe, and analyze any more than 50 responses would make MLR impractical as a clinical tool.

Two studies (Fisher, 1934; Minifie, Darley and Sherman, 1963) report the temporal reliability of the MLR. Temporal, or test-retest reliability, reveals the consistency of an individual's performance on the same test over a period of time. Both studies report that individual children did not appear to be very consistent in their language usage from day to day. They (Minifie, et al.), do report, however, that measures consisting of means of three 50-response samples (150 responses appear to have adequate reliability for most research purposes.

The validity of MLR has been questioned by many investigators. In an attempt to evaluate MLR by outside criteria, studies were designed by Sherman, Shriner and Silverman (1965), Shriner and Sherman (1967), and Shriner (1967). It was assumed in these studies that the impression language makes upon others might serve as the outside criterion. Psychological rating scales were used to evaluate listener's judgments. It was assumed that, if observer's ratings of language development were not in agreement with assessment of children's language as determined by present measures, then it could be concluded that measures such as MLR are neither useful nor valid for making such an assessment. MLR had a higher correlation with scale values than any other predictor variable. Therefore, they concluded that "if a single measure is to be used for assessment of language development, this one (MLR) would appear to be the most useful among those studied" (Sherman and Shriner, 1967).



In a related study (Shriner, 1967) four linear-multiple-regression analyses were used to determine the best composite of several language measures, including MLR, for predicting psychological scale values of language development for children of four different age categories. For the youngest age group (mean age, four years, seven months), a new measure, referred to as a modified length-complexity index, also was evaluated. The results revealed that as the mean age of the groups for analysis increased, MLR lost significance as a predictor; that is, little systematic relationship was observed between the criterion, scale values of language development, and MLR for children above the age of approximately five years. For children who were younger than five years, the best single predictor was the length-complexity index (LCI). This index did, however, correlate highly with both MLR and with Templin's method for deriving a structural complexity score (SCS). The SCS as a measure is based primarily on the adult model for correctness, with older children conforming more to this model.

The results of Shriner's (1967) study indicated that response length does not appear to be a significant indicator of expressive language for children who are approximately five years of age and older because of increased response variability. For older children it appears as if other factors such as those measured by the structural complexity score are beginning to play a more important role for assessing expressive language. It was also reported in Shriner's (1967) study that a modified length-complexity index was the best single predictor for children who were younger than 5 years of age.

The Length-Complexity Index (LCI) was synthesized by Shriner (1967), based on the research of Menyuk (1964), Bellugi (1964), and Cazden (1965). Menyuk noted that sentence complexity relates to more than sentence length; it is also a function of the ability to apply increasingly differentiated rules for generating sentences. If a child uses a rule to generate a sentence and then proceeds to conjoin two or possibly three similar sentences, the utterance would be obviously increasing in length; however, the utterance would not be increasing in complexity.

The scoring method for the LCI is derived from Shriner (1967), Bellugi (1964), Cazden (1965), Hurley (1967), and Miner (forthcoming). The following are examples of the scoring method used in this study, scored both as to MLR and LCI. Symbols used are N(noun), M(any modifier), A(article), P(plural inflection), Poss(possessive inflection), Aux(auxiliary), PrP(present participle), V(verb), PreV(pre-verb), Ng(negative), PP(past participle), and ?(question).

EXAMPLE	SYMBOLS	MLR	LCI
Kitties	N + P	1	2
My legs	M + N + P	2	3
The mother cat's tail	A+M+N+Poss + N	4	5
Big old dog's bones	M + M + N + Poss + N + P	4	7
He's going	N + Aux. + PrP	3	4
He'd walked home	N + Aux. + PP + N	4	5
I don't want it	N + Aux. + Ng + V + N	4	6
Is he going?	Aux. + N + PrP + ?	3	6

The complete rules for scoring the LCI will be found in Appendix A. The LCI as a language measurement tool is still very new, but future research should permit refinement of it. Nevertheless, the LCI does provide more information regarding the morphological and syntactical features of children's language than does either MLR or SCS.

While the LCI has been employed in single testing situation, its temporal reliability is still an unanswered but researchable question.

## CHAPTER III

### SUBJECTS, PROCEDURES, EQUIPMENT

#### Selection of Subjects

The seventeen subjects who participated in this study were children living in Sullivan, Illinois from September through December, 1967. They were selected on the basis of age, sex, intelligence, socioeconomic status, auditory acuity, physical status, and family language background. These criteria for the selection of subjects are discussed below.

A. Age and sex.--There were seven males and ten females selected on the basis of age and attendance at Kindergarten in the Sullivan Public Schools. The mean CA for the males was five years, one month, with a range from four years, eleven months to five years, three months, and a standard deviation of one and one-half months. The mean CA for the females was five years, two months, with a range from four years, ten months to five years, four months, with a standard deviation of two months. There was no statistically significant difference between the CAs of the males and females. The resulting student  $t$  ratio was .44 (df=15).

B. Intelligence.--Initially, only those children whose scores ranged from 80-130 on the Peabody Picture Vocabulary Test (Dunn, 1965) (a test of recognition vocabulary) were considered for inclusion in this study. This cutoff score was used because the IQ equivalents obtained on the Peabody tend to be systematically higher than the IQ scores on a test such as the Stanford-Binet (1951). The mean MA for the females was

five years, three months, with a standard deviation of eight months. The mean IQ for the females was 101.4, with a standard deviation of 7.92.

The mean MA for the males was six years, two months, with a standard deviation of eight months. The mean IQ for the males was 111.71, with a standard deviation of 5.40. There was a statistically significant difference in the MAs between males and females. The resulting  $t$  ratio of 2.55 (df=15) was significant beyond the .01 level. There was also a significant difference in the IQs between males and females. The resulting  $t$  ratio of 2.81 (df=15) was significant beyond the .01 level. There was a statistically significant difference between the CAs and the MAs of the males. The resulting  $t$  ratio of 3.90 (df=12) was significant beyond the .01 level.

C. Hearing.--A pure-tone audiometric sweep-check at 500, 1000, and 2000 Hz. in both ears was administered to all children. A Beltone audiometer, Model C calibrated to 1964 ISO standards, was used to present the pure tone at a level of 25 dB. re: .0002 dynes per  $\text{cm}^2$ . No children were eliminated for participation in this study on the basis of insufficient auditory sensitivity. All children included in this study were considered to have essentially normal hearing for the speech frequencies.

D. Physical status.--The subjects exhibited no obvious neuromuscular involvements as determined by observation of the experimenter.

E. Family language background.--No subjects came from homes with bilingual background. This particular community has a minority Amish population, and these children were excluded from this study. No subject was a twin, since previous research has shown that twins differ systematically



from the 'normal' language development pattern (Day, 1932; Davis, 1934).

Children who had a history of stuttering were also excluded from the study.

F. Socioeconomic status.—The families of the subjects were evaluated on the basis of the Index of Status Characteristics (Warner, 1949).

There are three separate steps in obtaining an Index of Status Characteristics for any individual or family. (1) Making the primary ratings on the status characteristics which are to comprise the Index—usually occupation, source of income, house type, and dwelling area. (2) Securing a weighted total of these ratings. (3) Conversion of this weighted total into a form indicating social-class equivalence.

There are seven occupational ratings ranging from 'professional and proprietors of large businesses' to 'unskilled workers.' There are seven sources of income ratings which range from 'inherited wealth' to 'public relief.' There are seven categories of house type, ranging from 'large houses in good condition' to 'all houses in bad condition—dwellings not meant for homes,' and a seven-point scale for rating dwelling areas.

The four categories are then multiplied by the weighting numbers 4,3,3,2 respectively. The weighted total may be any number from 12 to 84. If the ratings for any individual were all 1's, he would get a 12 (the highest rating in the upper class). If they were all 7's, he would rate an 84 (the lowest in the lower class). Children were randomly selected from a list of all five-year-old children attending in Sullivan, Illinois. It was assumed that such a procedure would result in cross-sectional sampling of the socioeconomic levels present in that community. It did not, however, as all subjects tended to fall at level two or below. Ten of the seventeen subjects were level five, lower class. One subject was upper

middle class, and the remaining six were lower middle and upper lower class.

### Procedures

Language samples were elicited from each child on three separate occasions. Half of the subjects were tested in the morning, and half were tested in the afternoon. Two examiners were utilized to elicit the verbalizations; each examiner saw the same child on each of the three testing situations. Both examiners were experienced in eliciting child language samples by virtue of their therapeutic and academic background.

The three speech samples were elicited within a ten-day period. The duration was short enough so that maturational influences on verbal output was assumed to be negligible. Sixty verbal responses were elicited from each child.

The experimenter constructed three sets of stimulus pictures judged by university speech pathologists to be of interest to five-year-olds. There was a random order of presentation between sets of pictures, but the order of presentation of pictures within a set was constant. These pictures were taken from pre-primers of several basic reading series. The pictures were presented to the child one at a time, and he was asked 'to tell a story about the picture.' In instances where the child had difficulty responding to the pictures, the examiners attempted to stimulate conversation by interjecting such questions as 'What else is happening in this picture?'

Each session was tape recorded on a Wollensak tape recorder, model T-1500. The experimenter listened to the tapes and transcribed the first 60 responses elicited from each child for each of the three recording sessions. The first ten verbalizations were discarded because they tend

to be shorter and less complex than later responses (McCarthy, 1930). Measures from analysis for each speech sample elicited included mean length of response (MLR), standard deviation of response length (SD-RL), number of total words (TNW), number of different words (NDW) and length-complexity index (LCI). The rules for computing the MLR have been defined by Templin (1957), following the McCarthy (1930) procedure. The rules for NDW and TNW or type-token ratio are defined by Miller (1951).

The TTR is the ratio of the number of different words (types) to the total number of words (tokens) in the passage. . . . One difficulty with the type-token ratio is that it gets smaller as the size of the sample gets bigger. If the passage contains only one word, this one word is one type and one token and so the TTR must equal 1.00. If the passage consists of two word tokens, these two tokens will probably be different types. We must take a passage of about 10 tokens before one type occurs more than once. As the length of the passage is increased it becomes more likely that words will be repeated and less likely that new, unused words will occur.

The rules for scoring the LCI were developed independently by Cazden (1965), Bellugi (1964), Hurley (1967) and Shriner (1967) and then synthesized by Miner (forthcoming).

### Recorder Reliability

Recorder reliability was estimated in order to determine the examiner's ability to identify spoken responses. What constitutes a verbal response is sometimes difficult to determine: (a) not all pauses represent breaks between responses; (b) some responses are made up of two short sentences; and (c) a single response may include a change in thought content. The McCarthy (1930) procedure was used to determine how well the examiner agreed with other observers in identifying spoken responses. The examiner and two other observers, professors in speech

pathology with experience in scoring child language samples, independently recorded the first 35 remarks of three five-year-old subjects. Each observer independently listened to the tape-recorded speech sample for three children and prepared a written script of what he considered to represent these 35 remarks uttered. The number of agreements and disagreements between the examiner and each of the other two observers were tabulated for each subject.

There was a 76% agreement among the experimenter and two independent observers relative to the transcription of the utterances of each of three subjects. Further analysis of this data revealed differences between the experimenter and the independent observers were due to different views on segmentation of utterances rather than differences in lexical items. These differences in segmentation were factored out, and the percentage of agreement was recomputed. The resulting agreement score was 99%. This was interpreted to mean that there was a high percentage of agreement between the experimenter and the independent recorders, except on the matter of segmentation. Therefore, it was necessary to re-examine the operational definition of what constitutes an utterance, both for MIR and LCI. For the MIR, the experimenter used the Templin (1957) definition of "per-breath utterance." For the LCI, the intent of the measure is to analyze a child's grammatical rules for his deep structure, not his surface structure. Sometimes an utterance will extend across a pause in order to complete the grammatical unit. Many times the sentence and the per-breath utterance will be the same language segment, but not always. The experimenter discussed the operational definitions of the



MLR and LCI with the two independent observers and all agreed that these were valid procedures to follow. Consequently, it was concluded that when the independent observers follow the same operational definitions for transcribing utterances, the utterances can be reliably transcribed from tape recordings as evidenced by the 99% agreement obtained.

### Scorer Reliability

Intrascorer agreement for the experimenter was found for two verbalization measures: MLR and LCI. From the typed speech samples the experimenter rescored 300 responses from six different subjects, for both Mean Length of Response and Length-Complexity Index. The obtained Pearson Product Moment Correlation Coefficient between the first and second scorings were 1.00, for both measures, showing the experimenter was satisfactorily consistent in scoring responses.

Inter-scorer agreement for the experimenter and one independent observer were obtained for both Mean Length of Response and Length-Complexity Index. Scoring of the LCI takes some orientation to complex rules, therefore only one observer was utilized. This independent observer rescored 200 responses from four different children. The resulting Pearson Product Moment Correlation Coefficient between the first and second scorings was 1.00, indicating perfect agreement between two examiners.

## CHAPTER IV

### RESULTS AND DISCUSSION

#### Results

The temporal reliability for the MLR and the LCI were computed by means of the Intraclass Correlation Coefficient (Ebel, 1951). The resulting correlations are indicated in Table 1.

TABLE 1

INTRACLASST CORRELATION COEFFICIENTS FOR MLR AND LCI

Language Measure	$r_1$	$r_{ave}$
MLR	.65	.85
LCI	.80	.92

The first correlation coefficient ( $r_1$ ) was computed to evaluate the reliability of an individual's responses on subsequent retests of single 50-response language samples. Inspection of ( $r_1$ ) in Table 1 reveals a higher temporal reliability for the LCI than the MLR. This is interpreted to mean that a composite linguistic analysis of length and complexity will yield a more consistent picture of an individual child's verbal output than will a measure of length alone.

The second coefficient ( $r_{ave}$ ) was computed to determine the group reliability of averages over three trials for both language measures.

The ( $r_{ave}$ ) differs from ( $r_1$ ) in that it is a measure of group consistency over three trials, whereas ( $r_1$ ) is a measure of individual consistency over three trials. Inspection of ( $r_{ave}$ ) in Table 1 reveals a slightly higher correlation for LCI than MLR. This is interpreted to mean that the temporal reliability correlation coefficients are higher for both measures on a group, as opposed to an individual basis. Clinically, the experimenter's greatest concern is with the consistency of an individual's linguistic performance, ( $r_1$ ), since language evaluations are concerned with individuals, not groups. In other words, it is important to know how representative a child's language sample is of his daily verbal performance over time. Consequently, Table 1 reveals that the LCI is a more consistent measure of verbal maturity than MLR.

Derived mean scores and standard deviations for MLR and LCI are reported in Table 2.

TABLE 2  
MLR AND LCI SCORES FOR EACH SAMPLE TIME

Language Measure	Time 1		Time 2		Time 3	
	M	SD	M	SD	M	SD
MLR	5.34	1.43	5.14	1.02	5.65	1.08
LCI	5.99	1.49	5.74	1.12	6.39	1.20

Inspection of Table 2 reveals that the derived MLR scores approximate very closely the data reported by Templin (1957, p. 79), for this age group and socioeconomic status. Furthermore the LCI mean scores are consistently

higher than MLR mean scores, because complexity as well as length is included in the scoring system. Rank ordering the mean scores for the three sample times, it should be noted that the highest mean score was obtained in Sample 3, followed by Sample 1, and then Sample 2. In an effort to assess the significance of the difference between the mean scores for the three sample times, student t tests were computed for both the MLR and the LCI. The resulting t ratios within both the MLR and LCI were nonsignificant, less than 1.00. This is interpreted to mean that within each language measure the derived mean scores were not significantly different and that this population was homogeneous in their responses from sample to sample.

A further analysis was made in order to determine the correlations within and between the two language measures. The resulting Pearson Product Moment Correlation Coefficients are reported in Table 3.

TABLE 3  
WITHIN AND BETWEEN CORRELATION COEFFICIENTS  
FOR MLR AND LCI

Measure	MLR-2	MLR-3	LCI-1	LCI-2	LCI-3
MLR-1	.78	.64	.99	.79	.65
MLR-2	--	.46	.78	.96	.48
MLR-3	--	--	.65	.49	.97
LCI-1	--	--	--	.79	.69
LCI-2	--	--	--	--	.53

\*r = .48 at 5% level, and .61 at 1% level (df=15)

Inspection of Table 3 reveals a wide range of correlations within and between the two language measures. For example, the Pearson  $r$  between MLR Sample 1 and LCI Sample 1 is .99. If one language sample only is taken, this might lead one to conclude that the correlation would always be high. Inspection of other single sample correlations is deceptive because of the range from .99 to .46. These ranges influence the interpretation of the reliability of the measures. The variation in correlations obtained seems to be related to Sample 3 for both MLR and LCI. Note that within MLR neither Sample 1 nor Sample 2 correlate highly with Sample 3. However, Samples 1 and 2 correlate highly with each other. The same pattern is revealed within the LCI samples. However, MLR-3 and LCI-3 are highly correlated suggesting that whatever variable(s) affected Sample 3, the effects are revealed in both measures. The variation in correlations for single sample responses reinforces the importance of determining the temporal reliability of any language measure. By this type of analysis, clinicians and researchers can assess whether measures of language development may be considered reliable.

### Discussion

There are several factors to consider in interpreting the data. One is the problem of examiner bias. There is the possibility that different examiners obtain different results, especially if examiner variables interact with subject and stimulus variables (Cowan et al., 1967).

Some intervening variable appears to have influenced the correlation with sample three since the obtained correlations for this sample time are the lowest observed. It is hypothesized that these low correlations are a



result of a time differential between the sample times. Examiner #1 elicited samples two and three on consecutive days, while Examiner #2 had a five day period between Samples two and three. Examiner #1 was ill during the taking of Sample three also, and the questions interjected by her were not of the same quality as during the other two sessions.

Another problem of interpretation is that of stimulus material bias. Very few studies have used the same stimulus material to elicit verbalizations. The sets of pictures used may not have been the most interesting to children of this age-group, despite the fact that they were so judged by the experimenter. This factor underscores the need to develop a standardized procedure for eliciting verbalizations in children. Media studies are now underway which should indicate which stimulus material among toys, still pictures, and movies produce the most verbalizations in children (Strandberg, Mintun, forthcoming).

A fourth variable to be considered in the interpretation of the data is that of sample size. There were only seventeen subjects available for analysis, all of whom met the criteria for participation in this study. The factor of sample size, for example, could account, in part, for differences between this study and the one reported by Minifie, Darley and Sherman (1963). They obtained intraclass correlation coefficient for the FMR ( $r_1$ ) = .82, in contrast to the ( $r_1$ ) = .65 in this study. Besides differences in sample size, it must also be remembered that their subjects were slightly older and represented different socioeconomic levels. While there were only seventeen subjects in the study, 2,550 sentences were analyzed. So the sample size is not as small as it might appear on first impression.

A fifth variable, which became evident after the samples were elicited, was the MA difference between the males and the females in the sample. The males had significantly higher MA's than the females, although there were only seven males and ten females.

The use of MLR in language analysis continues because McCarthy declared that no measure seems to have superseded it "for a reliable, easily determined, objective, quantitative, and easily understood measure of linguistic maturity" (McCarthy, 1954, p. 550). The fact remains, however, that after MLR is computed, all the clinician has is a numerical score for linguistic performance. It tells nothing about the grammatical structures a child has, or his ability to generate grammatical rules.

When the LCI is computed, its greater temporal reliability, as indicated in Table 1, shows it to be a more stable indicator of a child's verbal maturity than the MLR. The examiner has more information from its linguistic analysis. Computation of the LCI takes more skill and orientation on the part of the examiner, but the results justify the time spent. In addition, it has profound implications for planning therapy. For example, inspection of the noun phrase index (NPI) and the verb phrase index (VPI) should indicate the child's response strengths and weaknesses. If he has more VP's than NP's it might be wise to begin therapy with teaching the development of NP's. If a child has a rule for NP's which is  $A + N$ , and another rule which is  $M + N$ , the clinician may wish to begin teaching with the NP rules:  $A + M + N = NP$ . Other therapy approaches can be considered. If a child has a rule for

generating a noun, does he also use the rule for forming plurals or the possessive form of that noun? If he produces a verb, does he use both the present and past tenses? Does he use auxiliaries? Auxiliaries must be mastered before grammatically acceptable questions can be produced.

The LCI could be applied to other categories of language impairment. In aphasia, it would be important to know what linguistic rules exist following neurological damage. With skillful interpretation, the LCI approach should help to distinguish between those children who are delayed in language development, and those who are disordered in their language development.



## CHAPTER V

### SUMMARY AND CONCLUSIONS

#### Summary

The purpose of this study was to determine the test-retest reliability of a new measure of children's language, the Length-Complexity Index (LCI). For the LCI to be considered a reliable measure, it is necessary that several language samples taken over a period of time yield similar results.

A review of the literature revealed that Mean Length of Response (MLR) was the measure traditionally used to analyze the verbal output of children. MLR was developed by McCarthy (1930) and standardized by Templin (1957). Menyuk (1964) noted that sentence complexity relates to more than sentence length; it is also a function of the child's ability to apply increasingly differentiated rules for generating sentences. If a child uses a rule to generate a sentence and then proceeds to conjoin two or possibly three similar sentences, the utterance would be obviously increasing in length; however the utterance would not be increasing in complexity.

Shriner (1968), in an extensive review of the literature, questioned the adequacy of the MLR as a means of measuring developmental changes in language facility. A new measure of children's linguistic maturity, the Length-Complexity Index (LCI), was developed by Shriner (1967). The LCI

had its origin in the recent developments in structural linguistics (Shriner and Sherman, 1967). It purports to be a more sensitive measure of linguistic performance in children by analyzing length and complexity together. It was also reported in Shriner's (1967) study that a modified LCI was the best single predictor of language maturity for children who were five years of age and younger. The temporal reliability of this measure had not been determined until this time.

There were seventeen subjects who participated in this study, seven males and ten females. The subjects were five-year-old children attending Kindergarten in the Sullivan, Illinois Public Schools. Each of the subjects had normal hearing, average or above average intelligence as measured on the Peabody Picture Vocabulary Test (Dunn, 1965), no obvious neuromuscular disorders, and American English family language background. None were twins, or had a history of stuttering. All subjects came from middle or lower class socioeconomic level families.

Language samples were elicited from these children on three separate occasions within a ten-day period by two examiners. The stimulus material constructed to elicit the verbalizations consisted of three sets of pictures judged to be of interest to five-year-old children.

The language samples of three children were transcribed independently by the examiner and two professors in speech pathology. There was 76% agreement among the three judges on the transcription from tapes. There was a difference in the segmentation of utterances among the experimenter and the two independent recorders, rather than differences on lexical items. The differences in segmentation were eliminated, and the percentage of

agreement was recomputed. The resulting agreement score was 99% for all three recorders.

Intrascorer agreement for the MLR and LCI was found to have a correlation of  $r = 1.00$ . The experimenter re-analyzed 300 sentences for both the MLR and LCI. The resulting Pearson  $r$  between independent recorders was .99 for MLR. Since scoring of the LCI takes some orientation to complex rules, one independent scorer re-analyzed 200 sentences. The Pearson correlation between the independent scorer and the experimenter was 1.00, which indicated perfect agreement in the application of LCI scoring rules.

The intraclass correlation coefficient for MLR was  $r = .65$  for the individual child's responses on subsequent retests of single 50-response language samples. This indicates the considerable variability of MLR as a measure of a child's daily verbal output.

The intraclass correlation coefficient for the LCI was  $r = .80$  for the individual child's responses on subsequent retests of single 50-response language samples. The results indicate that as a language measure the LCI is not as variable as the MLR; that it tends to measure children's language output more reliably over time. This is interpreted to mean that a composite linguistic analysis of length and complexity will yield a more consistent picture of an individual child's verbal output than will a measure of length alone.

There are many implications for therapy using the LCI as a measurement tool. Both numeric and linguistic analyses of the LCI can be computed. The numeric procedure may be utilized as a pre- and post-test of developmental changes occurring as a result of maturation and therapy. This technique enables the clinician to quantify the qualitative aspects

of verbal maturity. Inspection of the noun phrase index (NPI) and the verb phrase index (VPI) should identify the response variability in a child's verbal output. If a child produced more NPs and VPs, the clinician might want to increase his client's verbal maturity by teaching the child to embed NPs in VP<sub>2</sub>. In this manner, the kitty can become I see the kitty. If more VPs than NPs are generated, it might be well to begin language therapy by teaching the development of NPs.

In linguistic analysis of the child's output, it would be important to identify those generative rules which are restricted to a child's grammar, those utterances which are considered ungrammatical by adult English standards, such as Me go. Furthermore, the list of NP and VP constructions can be viewed as a tentative hierarchy of developmental levels for that child. If a child has a rule for NP which is A + N, and another NP rule M + N, where the kitty becomes the yellow kitty. Knowledge of the child's environment and the relative frequency of occurrence of English words (Thorndike, 1944) should indicate which specific words to teach. Other therapy approaches should also be considered. If a child has a rule for generating a noun, does he also use the plural and possessive forms of that noun? If he produces a verb, does he use both the present and past tenses of that verb? Does he also use auxiliaries? Auxiliaries must be mastered before grammatically acceptable questions can be produced.

The LCI procedure can also be applied to other categories of language impairment. In mental retardation, the clinician should know which grammatical rules a child possesses before attempting to expand his verbal maturity. With skillful interpretation the LCI approach should

help distinguish between the child who is delayed in his language development and the child who is disordered in language development.

### Conclusions

From the results of the present study on its temporal reliability, the LCI would seem to be a more reliable measure of verbal output in children than the MLR. Furthermore, on the basis of previous research by Shriner (1967) it would seem to be the best single measure of a child's language abilities.

### Implications for further research

There are many implications for further research as a result of this study. The study should be replicated with a large number of subjects in the hope of establishing some norms for the LCI. It should be repeated using children of different age levels. If this were done, it would enable the therapist to compare a specific child with a large number of peers as Templin (1957) did with the MLR.

The weighting and scoring system of the structural complexity part of the LCI needs further research. Many questions have arisen concerning pronouns, possessives, levels of negatives, levels of questions, and irregular verbs. The sequence of emergence of verbs has not been established completely, and until it has, scoring of the past tenses of irregular verbs remains difficult.

There also needs to be a standardized procedure for eliciting verbal responses from children. At the present there is significant variability among examiners, stimulus materials and subjects (Cowan et al., 1967).



The size of the language sample necessary to compute a valid and reliable LCI is another unanswered, but researchable question. Until this issue is examined, the a priori assumption is to follow the lead of Darley and Moll (1960). They conclude that the average of three samples of 50 responses each would determine a child's "true" MLR, depending on the degree of precision needed by the researcher. There is no reasoning to support the inference that the MLR size of language samples should also prove to be meaningful for the LCI. It does seem plausible, however, to assume that the larger the sample, the more closely the LCI measure approaches the "true" mean for a specific child.

The present LCI was developed for analysis with a random selection of midwestern, white children. It presupposes that the generative rules for this particular dialect group and its sequential development is known. If the LCI is to be applied to a different dialect, the examiner must first identify what different rules for generating sentences exist in that dialect. Therefore, the propriety of generalizing the LCI to other dialects must still be examined.

## APPENDIX

### SCORING PROCEDURES FOR LCI

TRANSCRIBING THE RESPONSES. Record precisely, paying particular attention to inflected endings, pauses and repetitions. Mark off each incomplete or complete sentence (not per breath utterance as in MLR) with hash marks (/). Number each sentence consecutively beginning with number 1. In each sentence, underline NP<sub>1</sub> with a single line and VP<sub>2</sub> with a double line.

WORD COUNT. Subject and predicate contractions count as two words (same as MLR procedure). Note, some contractions occur in spoken English that are not considered grammatical in written English: it's, it'll, we're, we'll, that's, that'll, what's, what'll, you've, you'll, I'm, I'll, they're, they'll, she's, she'll, he's, he'll, who's, who'll, mine'll mine's, where's, where'll, I'd, you'd, he'd, she'd, it'd, they'd, we'd.

Contractions of the verb and negative are counted as one word: didn't, aren't, won't, don't, can't, ain't, wouldn't, couldn't, shouldn't, isn't.

Hyphenated words and compound nouns, particularly proper nouns designating a single object, are counted as single words: merry-go-round, cowboy, bubblegum, Miss X, doughnut, ABC's, jack-o-lantern, kool-aid, Santa Claus, Mother Goose.

Starters are eliminated and not scored: oh, and, then, now, um, hey, cause, well, Miss X. However, if any of these words serve a

sequencing function rather than as starters, they should be included and counted.

All prepositions are counted except in the following situations:

(A) when it is considered part of the infinitive construction: I'm ready to eat; I like to read. (B) When it is the last word in a sentence and is elliptical: He want to; I like to.

Omit word and/or phrase repetitions when (A) the same word is repeated several consecutive times, count the word only once. (B) When a phrase is repeated, count it only once unless one or more words is different; in that case, count only the phrase with the highest LCI point value. (C) If a word repetition occurs within a phrase repetition, count the word only once. (D) If a contraction is separated in a phrase repetition, count only the phrase repetition with the highest LCI score. (E) Repetitions for emphasis or constituting a dysfluency should be excluded.

Proper names in apposition are eliminated: Joseph, what are you doing? Mister, you got a flat tire.

NOUN PHRASE. Adjectives which are functioning as nouns are counted as residing in the noun phrase: Some more red; big fat two.

Pronouns serving in the nominative function are counted as noun phrases: I don't know what to do; I see it.

Noun phrases are not considered to extend across pauses.

Pauses frequently make structures ambiguous. Furthermore, Brown and Bellugi (1954) present a strong case for the psychological unity of the NP as a sentence constituent. In the following sentence, count only the underlined words: This is -- a dog.



N + N combinations are counted as single nouns on the NP index.

Score as one point: picture stove, telephone ball, tree bird, wrist watch, candy cane, department store.

Progressive pronouns are counted only if the correct term is used.

The intent here, according to Gazden (1965), is not to penalize for incorrectness, but to give credit only where the structure is clear:

Your shirt = 2 points; you shirt = 1 point.

A is not counted as an article when it is obviously a reduction of another word. It was considered a reduction of of in some a this and a reduction of it in take a back.

Plural inflections are not counted separately for a few words which are frequently utilized only as pluralized nouns: scissors, pants.

Most nouns form their plurals by adding /+s/, /+z/, or /+es/.

A few nouns change form: man, men; child, children. These should be considered appropriately as plural forms and scored as 2 points.

Noun phrase examples and assigned weights. Symbols: N(noun or pronoun), A(article), P(plural inflection), Poss(possessive inflection), Prp(preposition).

<u>Symbols</u>	<u>Examples</u>	<u>Score</u>
A	an, an, the	1
M	big, white, such	2
N	dog, dish	1
A+N	the dog	2
M+N	big dog	2
N+P	dogs	2
N+Poss	dog's	2
A+M+N	the big dog	3
A+N+P	the dogs	3
A+N+Poss	the dog's	3
M+N+P	big dogs	3
M+N+Poss	big dog's	3
Prp+A+N	by the dog	3

<u>Symbols</u>	<u>Examples</u>	<u>Score</u>
M+M+N	big white dog	4
A+M+N+P	the big dogs	4
A+N+Poss+N	the dog's dish	4
A+M+N+Poss	the white dog's	4
A+M+M+N	the big white dog	5
A+M+N+Poss+N	the big dog's dish	5
M+M+N+P	big white dogs	5
A+M+M+M+N	the great big old dog	6
A+M+M+N+Poss+N	a big old dog's dish	6
A+M+N+Poss+N+P	a big dog's dishes	6

**VERB PHRASE.** An unmarked verb is one lacking a suffix. Frequent suffix are +ed, +s, or +ing. Other verbs, even though they may be in the past tense, are considered unmarked verbs. The exception to this rule is the word said. Said is considered to be a regular past tense verb. The rationale for this is that the word is in effect the same as sayed and thus is considered to possess a suffix.

In infinitive constructions, the word to is considered to be part of the verb and not a preposition. Thus, the word to in this case is not scored. Furthermore, the word to, since it is an elliptical expression standing for an infinitive, is not scored.

Only lexical verbs and connectives are counted. This procedure eliminates the problem of deciding when particular prepositions are considered part of the verb and when they are not, especially for cases other than the infinitive. For example, in the sentence, "Pick out mine," the question of whether the verb is pick or pick out would depend on such factors as intonation, normal usage of the expression by the child and other considerations not determinable through a tapescript. One notable exception exists relative to the rule of counting only lexical verbs. Preverbs are frequently observed in the verbal output of children. Since

they indicate the transitional development of a grammatical rule for verb forms, credit for this performance should be given. Score all preverbs as 1 point: goms, oughta, shoulda, coulda, woulda, and halfta.

Since pauses always contribute some ambiguity, a verb is counted only if it is on the same side of the pause as its subject. In Monny want me--put this on, want receives a score of 1; similarly wanna, in Her wanna--hold this.

The verbs in each phrase are counted separately. In You saw we had turkey, saw and had each receive 1 point.

In the case of a compound predicate, both verbs are counted if they receive the same score; if not, only the verb closest to the subject is counted. In Somebody jumps and bites, each verb receives 2 points. In He's coming and get out, only the first verb is counted, for a score of 3. This rule prevents any penalty for a correct usage of allipsis.

No penalty is computed for errors. Only correct responses or obvious approximations are tabulated. The verb phrase weights for some unique constructions are indicated as follows: I doed, it broked. Each verb is scored as two points (verb+Past Tense). He's upped = 3 (aux+V+Pst).

Scoring of verbs presents many complex and subtle problems. Regular verbs usually form the past tense by adding +ed: jump-jumped, look-looked. Each past tense suffix receives one point. Irregular verbs indicate tense differently: run-ran, come-came, think-thought. Score all irregular past tense verbs as 1 point. This may seem incongruous in some situations since doed receives two points and did receives only 1 point. This paradox can only be resolved after more is known about the emergence of verb forms in children. Given this information,

one should be able to assign weights to the three forms of the past tense (past, perfect and past perfect) and two forms of the future tense (future and future perfect). The most important consideration until more definitive developmental information exists is that the clinician be consistent in his scoring procedure.

Verb phrase examples and assigned weights. Symbols: V(verb), PrPt(present participle), Aux(auxiliary), P(plural), PsT(past tense), PreV(preverb), PP(Past participle).

<u>Symbols</u>	<u>Examples</u>	<u>Score</u>
PreV	gonna	1
V	go, is, jump	1
V+P	goes, jumps	2
PrPt	going, jumping	2
Aux+PrPt	is going	3
Aux+PP	had jumped	3
Aux+V	can jump	2
Aux+PreV+V	is gonna go	3
Aux+Aux+V+PsT	could have gone	4
V+V	try to go	2
Aux+PrPt+V	am going to get	4
Aux+PP	have arrived	3
Aux+Aux+V+PrPt	could have been going	5
Aux+PrPt+V+V	am going to try to fix	5

Negatives. The following point system for negatives and questions was based on the research of Ballugi (1966). Four different point levels are operationally defined as regards the usage of negatives.

The negation appears either at the beginning or at the end of the utterance, not within, and consists of no or not and the rest of the sentence. Score as 1 point: no wash; no singing song; wear mitten no.

Two auxiliary verbs appear in the negative form, can't and don't. The negative element now appears within the sentence but is not yet connected to an auxiliary verb. Score as 2 points: nominal+no, can't,

don't+main verb. Examples: I no bite you; I can't catch you; I don't want it. Furthermore, at this point level, the negative also appears in the demonstrative form at the beginning of a sentence in the imperative form. Demonstrative+no or not+nominal: That no mommy; that no fish school. Also observed is don't+main verb: Don't leave me.

When the negative form appears between the noun phrase and the present participle, assign a weighting value of 3 points. NP+Ng+PrPt: Me not crying; I no peeking.

The last level exemplifies the adult version of the negative. The sentence includes appropriate intonation. Score as 4 points: No, it isn't or No, I don't have the book. Auxiliaries are contracted with the negative n't: You didn't eat supper with us; I can't see it. These sentences are of the form: Nominal+Aux+Ng+V. In child language the verb be is often missing but is now optional. Nominal+(be)+not+nominal objective: That not a clown or I am not a doctor.

QUESTIONS. Questions are formed primarily by a rising intonation, with and without a wh word. Bellugi (1966) distinguishes two levels of questions. For the first level, there are no auxiliaries and no subject-verb inversion. There are few negative questions. Score as 1 point: Mommy eggnog? I ride train? Who dat? No ear? What cowboy doing?

At the second level, yes-no questions contain an auxiliary or some form of do. Score as 2 points: Aux+nominal+V+?; Is mommy talking? Did I hit? The auxiliary component can have an optional negative attachment. Aux+Ng+nominal+V+?; Can't you work? Sometimes the auxiliaries are not inverted: What he can ride in? Why the kitty can't stand up? The auxiliary is optional in wh questions: What is he writing? What he is writing? What he writing?



## BIBLIOGRAPHY

- Bellugi, Ursula. The Emergence of Inflections and Negation Systems in the Speech of Two Children, paper presented at New England Psych. Assn., 1964.
- Braine, M. "The Ontogeny of English Phrase Structure: the First Phase." Language, 39, 1963, 1-13.
- Brown, R., and Bellugi, U. "Three Processes in the Child's Acquisition of Syntax." Harvard Educ. Rev., 34, 1964, 133-151.
- Brown, R., and Fraser, C. "The Acquisition of Syntax" in Verbal Behavior and Learning. C. M. Cofer and B. S. Musgrave (eds.). New York: McGraw-Hill, 1963.
- Carroll, J. B. "Language development in Children" in Psycholinguistics: a Book of Readings. S. Saporta (ed.). New York: Holt, Rinehart and Winston, 1961, 334.
- Cazden, Courtney B. "Environmental Assistance to the Child's Acquisition of Grammar." Unpublished doctoral dissertation, Harvard University, 1965.
- Chomsky, N. Aspects of the Theory of Syntax. Cambridge, Mass.: MIT Press, 1965.
- Cowan, P.A., Weber, J., Hoddinott, B.A. and Klein J. "Mean length of Spoken Response as a Function of Stimulus, Experimenter, and Subject." Child Development, 38, 1967, 191-203.
- Darley, F.L., and Moll, K.L. "Reliability of Language Measures and Size of Language Sample." J. Speech Hearing Res., 3, 1960, 166-173.
- Davis, Edith. "Mean Sentence Length Compared with Long and Short Sentences as a Reliable Measure of Language Development." Child Development, 8, 1937, 69-79.
- Day, Ella J. "The Development of Language in Twins; I. A Comparison of Twins and Single Children." Child Development, 3, 1932, 179-199.
- Dunn, L.M. Manual for the Peabody Picture Vocabulary Test. Minneapolis: American Guidance Service, 1965.

- Ervin, Susan. "Imitation and Structural Change in Children's Language" in New Directions in the Study of Language. E. Lenneberg (ed.). Cambridge: MIT Press, 1964.
- Fraser, C., Bellugi, U., and Brown, R. "Control of Grammar in Imitation, Comprehension and Production." J. Verb. Learning, Verb. Behavior, 2, 1963, 121-135.
- Lee, Laura, L. "Developmental Sentence Types: A Method for Comparing Normal and Deviant Syntactic Development." J. Speech Hearing Dis., 31, 1966, 311-330.
- McCarthy, Dorothea A. "Language Development in Children," in L. Carmichael (ed.). Manual of Child Psychology. New York: Wiley and Sons, 1954, ch. 9.
- McCarthy, Dorothea A. "The Language Development of the Preschool Child." Inst. Child Welf., Monogr. Ser., No. 4, Minneapolis: University of Minn. Press, 1930.
- McNeill, D. "The Capacity for Language Acquisition." Volta Rev., 68, 1966, 17-33.
- McNeill, D. "Developmental Psycholinguistics," in F. Smith and G.A. Miller (eds.), The Genesis of Language, Cambridge: MIT Press, 1968, 15-84.
- Menyuk, Paula. "Syntactic Structures in the Language of Children." Child Development, 34, 1963a, 407-422.
- Menyuk, Paula. "A Preliminary Evaluation of Grammatical Capacity in Children." J. Verb. Learning, Verb. Behav., 2, 1963b, 429-439.
- Menyuk, Paula. "Syntactic Rules Used by Children from Preschool Through First Grade." Child Development, 35, 1964a, 533-546.
- Menyuk, Paula. "Alteration of Rules in Children's Grammar." J. Verb. Learning, Verb. Behav., 3, 1964b, 480-488.
- Menyuk, Paula. "Comparison of Grammar of Children with Functionally Deviant and Normal Speech." J. Speech Hearing Res., 7, 1964c, 109-121.
- Miller, W., and Ervin, S. "The Development of Grammar in Child Language, in The Acquisition of Language. U. Bellugi and R. Brown (eds.). Monogr. Soc. Res. Child Development, 28, 1964, 9-34.
- Minifie, R., Darley, F.L., and Sherman, Dorothy. "Temporal Reliability of Seven Language Measures." J. Speech Hearing Res., 6, 1963, 139-149.

- Mintun, Shirley. "A Preliminary Investigation of Certain Procedures for Eliciting Verbalizations in EHH Children." Unpublished Master's Thesis, Eastern Illinois University, 1968.
- Fovich, Edna and Baratz, Joan. "Grammatical Constructions and the Language of the Negro Preschool Child," paper presented at American Speech and Hearing Association, Chicago, 1967.
- Revised Stanford-Binet Intelligence Scale, Third Revision. Boston: Houghton-Mifflin, 1951.
- Sherman, Dorothy, Shriner, T.H., and Silverman, F. "Psychological Scaling of Language Development of Children." Proc. Iowa Acad. Sci., 72, 1967.
- Shriner, T.H. "A Comparison of Selected Measures with Psychological Scale Values of Language Development." J. Speech Hearing Res., 10, 1967a, 828-835.
- Shriner, T.H., and Sherman, Dorothy. "An Equation for Assessing Language Development." J. Speech Hearing Res., 10, 1967b, 41-48.
- Shriner, T.H. "A Review of Mean Length of Response as a Measure of Expressive Language Development in Children." J. Speech Hearing Dis., (in press).
- Siegel, G.M. "Inter-examiner Reliability for Mean Length of Response." J. Speech Hearing Res., 5, 1962, 91-95.
- Strandberg, Twila. "A Preliminary Investigation of Certain Procedures for Eliciting Verbalizations in Normal Children." Unpublished Master's Thesis, Eastern Illinois University, 1968.
- Templin, Mildred O. "Certain Language Skills in Children, Their Development and Inter-relationships." Inst. Child Welf. Monogr. Ser., No. 26, Minneapolis: Univ. of Minn. Press, 1957.
- Thomas, O. Transformational Grammar and the Teacher of English. New York: Holt, Rinehart, Winston, 1966.
- Thorndike, L.E. Teacher's Wordbook of 30,000 Words. Bureau of Pub. Teachers Coll., Columbia Univ., 1944.
- Warner, W.L., Meeker, M., and Eells, K. Social Class in America (a manual of procedure for the measurement of social status). Chicago: Science Research Assoc., 1949.